Date of acceptance Grade

Instructor

Interpretation of genetic interactions

Sirkku Karinen

Helsinki April 4, 2007

UNIVERSITY OF HELSINKI
Department of Computer Science

Contents

1 Introduction

Because rather simple model organisms have contributed in our knowledge of function of human genes, it is assumed that model organisms can be utilized on research of genetic interactions too [GHH01]. Yeast is an important model organism when we want to achieve knowledge of conserved biochemical processes [FiT03]. Also information about yest genes working together may shed some light on human genetic variation.

Sequence based comparison between species is already common procedure [ShI06]. Next step is to compare biological networks. This kind of comparison could be used to predict new protein functions and maybe this knowledge will tell us something more about evolution of proteins and species. Because different kind of biological networks concentrate on different aspects of networks it might be useful to combine these different data sources to get new picture about interactions within a cell.

If two genes together causes lethal phenotype, this genetic interaction is called synthetic-lethal [GHH01]. Physical interpretation of synthetic-lethal genetic interactions could reveal the functional meaning behind these genetic interactions [IdK05]. Synthetic-lethal interactions can be extracted from yeast quite easily.

Mapping of synthetic-lethal interactions in yeast is fast process because methods like SGA and SLAM have automated it [IdK05]. But determining functional significance of these interactions is still very slow. Ideker and Kelley [IdK05] suggest that these synthetic-lethal interactions could be combined with physical interactions in order to interpret functionality behind synthetic-lethal interactions. They have built a framework that assembly genetic interactions and physical interactions of yeast into models.

Basis of this seminar report is Ideker and Kelleys paper **Systematic interpretation of genetic interactions using protein networks** [IdK05]. At the end of this report I represent their framework. First chapters describe synthetic-lethal interactions and models generalized from these interactions.

2 Synthetic-lethal genetic interactions

In many processes one defect doesn't effect on the outcome of the process. Only when defects cumulate on some functionality, whole process will fail. For instance

if one rung in your ladder is broken, you probably still can use the ladder. If you brake another rung beside the first broken one, ladder becomes useless. This analogy applies on genetic background of variation in phenotype [GHH01]. If one gene is deleted cell may function correctly if some other gene is still working fine. Only after both genes are not working, cell expresses lethal phenotype. If mutations in two different proteins cause a disease, this relation is called synthetic-lethal genetic interaction. In many cases one gene alone doesn't affect on phenotype because functionality of genes is buffered with other genes. In interaction network of cell there seems to exist buffering in genetic variation [GHH01]. One gene may buffer variation in an another gene. Identical mutation may produce different phenotypes in different individuals. If gene A buffers variation of gene B, there is at least one allele of gene A that causes gene A to lose it's capability to buffer variation in gene B. This buffering may cause synthetic-lethal relation of two genes if gene A buffers otherwise lethal variation in gene B.

There already are methods for detecting these synthetic-lethal interactions in yeast automatically. One method is synthetic genetic arrays (SGA) [Ton01]. In this method there is an array with approximately 4700 plates that each contain different yeast knockout. Cells on each plate are still viable. Then studied query mutation is inserted to each plate. If cells on the plate stop growing or they die, combination of knockout originally on the plate and inserted mutation is synthetic-lethal or synthetic-sick. Growth of mutants is monitored with automated image analysis.

Other method is synthetic lethal analysis of microarrays (SLAM) [OSB03]. This method is similar to SGA, but mutants are grown in pools. In one pool there are only these 4700 viable knockouts and in the another pool there are the same knockouts with query mutation. Every deletion has unique sequence flanking and this can be used in analysis. After cells are grown in the pools controls and mutants are hybridized in a microarray and differences in intensities describe which mutants have grown and which have died.

3 Physical interpretations

Synthetic-lethal interactions has been mapped into three kind of interpretations: between pathway-models, within-pathway models [GHH01], [IdK05] and indirect effects [FiT03].

Between-pathway model describes process where two genes in different pathways

conducts complementary or redundant tasks. These tasks may be biochemically distinct but interpreted functionally, tasks are the same. One example of between-pathway interpretations is DNA repair. There are several mechanisms how DNA is repaired and mutation at the same time in different mechanisms causes DNA repair to fail. In the figure 1 is represented between-pathway interpretation of genetic interactions. In this figure, there are physical pathways that are connected with several genetic interactions.

Within-pathway models are derived genes working in a same pathway or process. In the figure 2 is represented within-pathway interpretation of genetic interactions. In this model genetic interactions occur within a specific pathway. Although these interactions seem to be majority of synthetic-lethal interactions [GHH01] in Ideker and Kelleys experiment [IdK05] genetic interactions were assigned into between-pathway models three and half times more often than into within-pathway models. Dataset used in their experiment might be biased because SGA experiments are conducted only to genes that are not previously found to be essential for cell survival.

Indirect effects can not be mapped into a physical network. A cell may respond to mutation in a gene and that way it can affect to many different pathways causing synthetic lethal interaction. These kind of synthetic-lethal interactions are predicted to be rare. At experiment of Ideker and Kelley [IdK05] they noticed that they could interpret 40% of genetic interactions into between- or within-pathway models. Their method could not classify 60% of genetic interactions into either one of these models.

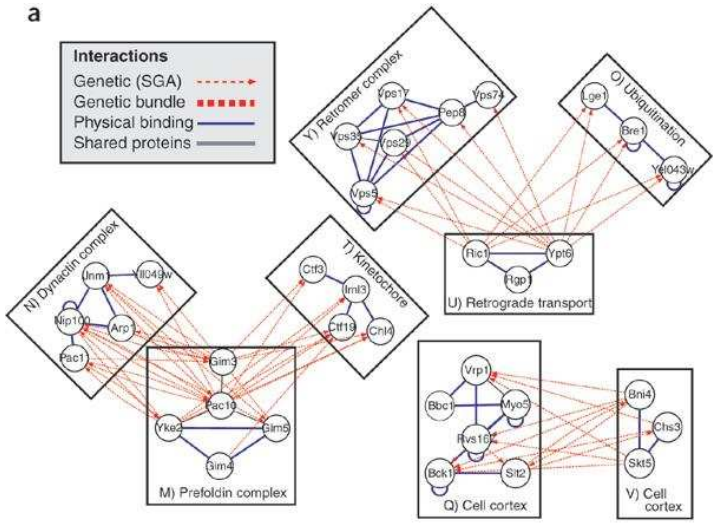

Figure 1: Between-pathway interpretations [IdK05].

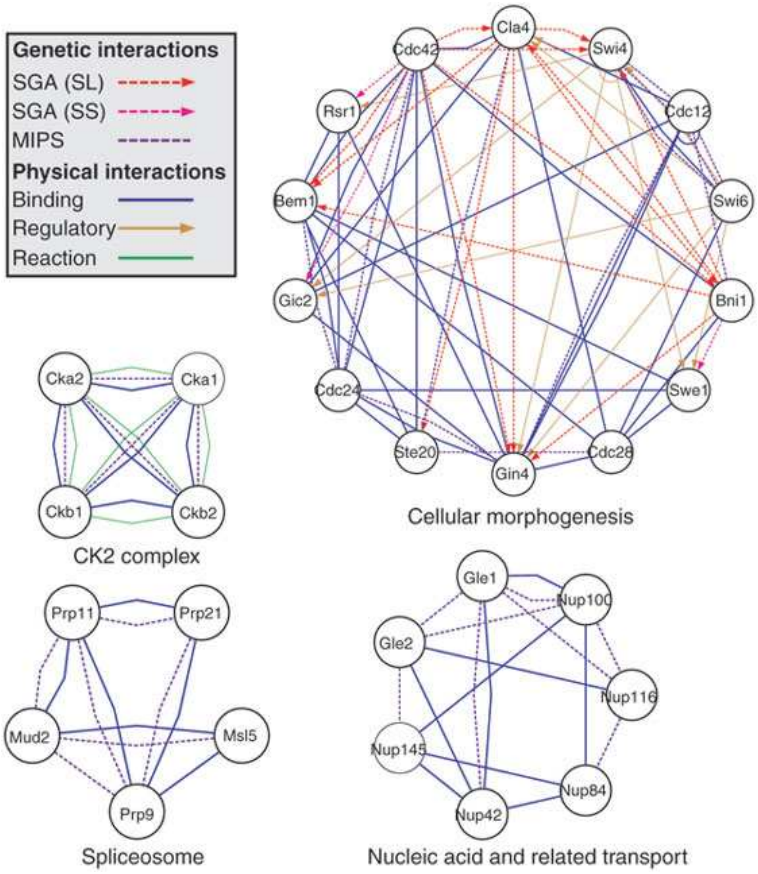

Figure 2: Within-pathway interpretations [IdK05].

4 Framework for finding interpretations

Although synthetic-lethal interactions can be searched automatically, functional interpretation of these interactions is very slow [IdK05]. Ideker and Kelley [IdK05] have demonstrated a systematic method to map genetic interactions on physical interpretation. They have built a framework that assembly genetic interactions and physical interactions into models generalized from physical interpretations of genetic interactions. Physical interpretations they used were between-pathway and within-pathway models. In this chapter I present the framework they have created and the results they got from testing their system.

Data they used in their experiment was gathered from different sources. 2012 synthetic-lethal and 2113 synthetic-sic interactions collected from SGA screening where 132 yest genes were deleted. Another 687 synthetic-lethal interactions was from Munich Information Center for Protein Sequences (MIPS). Eventually they had 1424 proteins linked with each other by 4812 synthetic-lethal interactions.

Data for physical networks was collected from DIP-database (protein-protein interactions), KEGG-database (enzymatic interactions) and from large scale study of 106 transcription factors (protein-DNA interactions). At the end physical networks covered 94,4% of genes in synthetic-lethal interaction dataset.

Steps in the experiment were first to map genetic interactions into physical interactions. Next physical interaction models were enriched by data about functional annotations from Gene Ontology database. After these steps new protein functions could be predicted from proteins in physical interaction models. Last step was to predict new genetic interactions. Next I describe more accurately each of these steps in framework of Ideker and Kelley.

4.1 Genetic interactions into physical interaction models

In the framework genetic interactions were mapped on between-pathway model and within-pathway model. Genetic interactions related to physical interactions only in limited cases. In this model each pathway of physical interaction networks included protein complexes and other network structures where set of proteins are densely connected by physical interactions. Interesting pathways were those pathways that had set of proteins that have denser genetic interactions than would expect in random. These pathways were extracted from physical models computationally.

Relation of genetic interactions to between-pathway models was constructed by probabilistic model. If genetic interactions is interpreted as between-pathway model, there is a pair of physical pathways that have dense genetic interactions in between. Pathway pairs were constructed if there were connected with many genetic interactions. All found pairs was scored according to their density of connective genetic interactions and density of physical interactions within pathway. Drawback of this method is the fact that all datasets are not as predictive than others. Large networks are more likely to generate high scores randomly. Comparison with random genetic interaction networks was conducted in order to determine significance of the models.

Within-pathway interpretations of genetic interactions fall into pathways that have, beside of physical interactions, also dense genetic interactions. For within-pathway model scoring was different. Scoring captured group of proteins that were interacting with more than would happen in random. Model for this gave higher scores on set of proteins that were interacting by both genetic interactions and physical interactions.

4.2 Functional annotations for models

For validating the models functional annotations were included into them. Annotations were retrieved from Gene Ontology database. Proteins that had common molecular function in pathways were enriched with annotations. Functional role of proteins in a pathway had to be over significant level of P=0.05 in order to annotation to be added into model. Same enrichment was done to between- and within-pathway models.

4.3 Prediction of functions and interactions

After models were finished, new protein functions and genetic interactions were predicted from functionally annotated models. For physical pathways that most of their proteins had common functional annotation, rest of proteins were predicted to have same function. This method succeeded 63% for between-pathway models and 69% for within-pathway models in a cross validation test. In the test 20% of annotations were removed and predicted again with remaining annotations. Prediction was scored to succeed or fail.

In between-pathway models proteins in one pathway interact wit same partners in another pathway. This causes complete bipartite motifs to occur in genetic interaction network. In this motif two interacting proteins have every possible link into another two interacting protein. If motif is not complete and one link out of four is missing, this implies that missing interaction is also true. These predictions were also validated with cross validation. In eight incomplete motifs this method predicted correctly 87% of genetic interactions. This method relies on between-pathway model and if these incomplete motifs were searched from all models, prediction accuracy fell to 5%.

In within-pathway models genetic interactions were predicted to proteins that had one or more common neighbors. Cross validation test revealed that best prediction accuracy of 38% was reached when threshold of number of common neighbors was set to three. If predictions weren't made only within-pathway model, corresponding prediction accuracy was only 15%.

Physical interpretations of genetic interactions had a major impact on prediction accuracy.

5 Conclusions

Some genes are not alone essential to individuals genotype. They interact and some combinations of variants of these genes are lethal. These synthetic-lethal genetic interactions can be quite easily found from yest with high throughput methods like SGA and SLAM. Interpreting physical interactions have been more laborious task. Information of these genetic interactions can be combined with interaction networks and functional knowledge of proteins in these networks. This approach gives some physical explanation for genetic interactions and it can be used for predicting new protein functions and genetic interactions.

Protein functions and genetic interactions in yeast can help understanding same events in human and other species too. When biological networks in different species are compared with each other, it can be used to detect conserved networks and protein functions.

Ideker and Kelley [IdK05] present a framework that can be used on systematic search of physical relations behind genetic interactions. Assembly of data from various sources can be used for predicting new protein functions and genetic interactions.

References

GHH01 Garvik, B., Hartman, J., Hartwell, L., Principles for the Buffering of Genetic Variation. Science 9, 2001.

IdK05 Ideker, T., Kelley, R., Systematic interpretation of genetic interactions using protein networks. Nature Biotechnology 23, 2005.

FiT03 Fields, S., Tucker, C., Lethal combinations. Nature Genetics 35, 2003.

Ton01 Tong, A., et al, Systematic Genetic Analysis with Ordered Arrays of Yeast Deletion Mutants. Science 14, 2001.

OSB03 Ooi, S.,Shoemaker, D., Boeke, J., DNA helicase gene interaction network defined using synthetic lethality analyzed by microarray. Nature Genetics 35, 2003.

ShI06 Sharan, R., Ideker, T., Modeling cellular machinery through biological network comparison. Nature Biotechnology 24, 2006.